

# **SUMMARY REPORT OF STANDARD DATA ELEMENTS FOR CHEMICAL SUBSTANCES**

**CONTRACT NO. 68-WI-0055  
DELIVERY ORDER NO. 57**

**Prepared for:**

**United States Environmental Protection Agency  
Office of Information Resources Management  
401 M Street, SW.  
Washington, DC 20460**

**Delivery Order Project Officer:**

**Geoffrey Steele**

**Prepared by:**

**EPA Systems Development Center  
(A Contractor Operated Facility)  
Science Applications International Corporation  
200 North Glebe, Suite 300  
Arlington, VA 22203**

## CONTENTS

<b>1.0</b>	<b>INTRODUCTION</b>	1
1.1	Background	1
1.2	Purpose	2
1.3	Scope	2
1.4	References	2
<b>2.0</b>	<b>METHODOLOGY</b>	3
2.1	Description of Proposed Standard Data Elements for Chemical Substances	4
<b>APPENDIX A</b>	Proposed Standard Data Elements for Chemical Substances	

## **1.0 INTRODUCTION**

The Environmental Protection Agency (EPA), Office of Information Resources Management (OIRM), Enterprise Information Management Division (EIMD), through the mission of its Information and Data Management (IDM) Program, has introduced the concept of centralized management and coordination of EPA's distributed information and data resources. The data policies and standards of the IDM program will be supported by data architecture, data models, standard data element domains, and a repository of standard data elements. These work products will facilitate data sharing efforts and effective management of EPA's Information Resources Management (IRM) resources. The focus of this report is the data elements and domains for proposed standard data elements for chemical substances.

### **1.1 Background**

The One Stop Reporting Program was initiated by the EPA Executive Steering Committee (ESC) for the purposes of improving the Agency's management of environmental information and reducing the burden of reporting across the regulated community. To achieve the purpose of improving information management the One Stop team has involved the stakeholders, particularly the states, in identifying the need for data standards and participating in their development. To further the goals of the One Stop Program, the Information and Data Management Service Center (IDMSC), Delivery Order (DO) 57, has been tasked to identify needed data standards that will facilitate data sharing and to develop these data standards.

In addition to the need to share information between the EPA and the One Stop States, the Reinventing Environmental Information (REI) Initiative recently undertaken by the Agency has the implementation of data standards and the development of electronic reporting processes as major goals. One of the core areas for implementation of data standards is the identification of chemicals. Previous efforts to integrate chemical data have been directed towards developing cross-references for the same and similar chemical substances. For example, the Register of Lists (ROL) is a tool for determining which chemicals are regulated by the EPA and how they are regulated. The Envirofacts Master Chemical Integrator (EMCI) is another step toward developing a master list of chemical names and common names which identify substances of interest to various Agency programs. EMCI provides links, based on chemical identity, to information in the following Envirofacts systems: Permit Compliance System (PCS), AIRS Facility Subsystem (AFS), RCRA Information System (RCRIS), and Toxic Release Inventory System (TRIS). Other efforts to utilize chemical information for program needs can be found throughout the Agency.

In the past the management and utilization of chemical information has been done within individual programs with systems developed to meet their specific needs. As a result, the Agency has not built a central source for systematic names for chemicals, prepared a

comprehensive list of common names, trade names, or synonyms, developed common identification numbers for chemical substances, or created a central source for quality chemical data. Because of these differences, the ability to easily exchange information within the Agency has been hampered, and the capability of those outside the Agency to access information has been limited. The development of standard data elements for chemical substances should contribute to the minimization of many of the problems that have been seen in the past.

OIRM currently has one data standard for identification of chemical substances. The *EPA Order - Chemical Abstract Service Registry Number Data Standard* (EPA Directive Number 2180.1) “establishes the policy and responsibilities related to the use of registry data, specifically the Registry Number, for Chemical Abstracts Service (CAS) Division of the American Chemical Society in automated information systems containing data/information on specific, definable chemical substances. The CAS registry number provides consistent and unambiguous identification of chemicals and facilitates sharing chemical information across programmatic media.” This data standard was approved in June 1987 and “applies to all EPA organizations and their employees as well as the facilities and personnel of agents of the EPA who design, develop, operate, or maintain Agency information and information systems having to do with specific, definable chemical substances.”

## **1.2 Purpose**

The purpose of this report is to propose standard data elements for chemical substances and to specify domains for those data elements that comply with EPA, national, and international standards.

## **1.3 Scope**

This document is limited to proposing data elements that can be utilized to identify chemical substances. These data elements represent information that has been recorded in numerous EPA information systems to identify chemical substances and the corresponding domains for those data elements that have been collated from values currently maintained in those information systems.

## **1.4 References**

The following documents were used in the preparation of the proposed chemical substances standard data domain values:

- *EPA - Chemical Abstract Service Registry Number Data Standard*, IRM Policy, Standards and Guidance/ IRM Strategic Planning Documents (EPA Directive No. 2180.1), dated June 26, 1987.
- *ChemID* [Online]. (1997). Available: National Library of Medicine: telnet: medlars.nlm.nih.gov.
- *Chemline* [Online]. (1997). Available: National Library of Medicine: telnet: medlars.nlm.nih.gov.
- *CSCHEMfinder* [Online]. (November 1997). Available: CambridgeSoft Corporation: <http://chemfinder.camsoft.com>.
- *Dictionary of Chemical Names and Synonyms*, Howard, P. And Neal, M. Lewis Publishers. 1992.
- *Envirofacts* [Online]. (November 1997). Envirofacts Master Chemical Integrator (EMCI), United States Environmental Protection Agency (USEPA). Available: [http://www.epa.gov/enviro/html/emci/emci\\_overview.html](http://www.epa.gov/enviro/html/emci/emci_overview.html)
- *The Merck Index, An Encyclopedia of Chemicals, Drugs, and Biologicals*, Susan Budavari, Editor. Merck & Company, Inc. 1989.
- *Register of Lists*, USEPA. (Agency LAN).
- *STNEasy* [Online]. (November 1997). Available: Chemical Abstracts Service: <http://stneasy.cas.org>.

## 2.0 METHODOLOGY

Data standards include two components: standard data elements recorded in the Environmental Data Registry (EDR) and standard data values for all data elements where only a discrete set of data values are acceptable. The data standards team analyzed how chemical substances are identified in the Envirofacts systems and developed the proposed data elements for chemical substances, which are listed in Section 2.1. The proposed standard data elements and domain values for standard data elements were then recorded in the EDR. The domain values currently recorded for the data elements “Chemical Abstracts Service Registry Number” and “Chemical Abstracts Service Registry Number Display Text” are the result of the intersection of Chemical Abstracts Registry Numbers in the ROL and EMCI. These values are representative of the type of domain information that will be recorded in other enumerated data

elements.

## **2.1 Description of Proposed Standard Data Elements for Chemical Substances**

In development of the proposed standard data elements for Chemical Substances, twelve individual data elements were defined. Detailed information about these data elements, their domains, and permissible values are provided in Appendix A, as reports generated by the EDR.

The proposed data elements for identification of Chemical Substances include:

- Chemical Abstracts Service Registry Number - The unique number assigned by Chemical Abstracts Service (CAS) to a chemical substance. (DE 6165, 9-numeric, mandatory)
- Chemical Abstracts Service Registry Number Display Text - The output display text that represents the Registry Number assigned by Chemical Abstracts Service to a chemical substance (CASRN). (DE 5893, 11-character)
- Chemical Substance Index Name - The name assigned to a chemical substance by Chemical Abstracts Service (CAS) that provides information about its molecular composition. (DE 5799, 240-character)
- Chemical Substance Common Name - The generic name that is commonly used to identify chemical substances. (DE 5801, 240-character)
- Chemical Substance Synonymous Name - Alternate names that are commonly used to identify a chemical substance. (DE 5806, 240-character)
- Chemical Substance Trade Name - The name of a commercial product that contains the chemical substance as an active ingredient. (DE 6103, 240-character)
- Chemical Substance Type Name - A descriptive name that records composition information about a chemical substance that is not clearly identifiable as a single, specific chemical compound. (DE 5818, 20-character)
- Chemical Substance Type Code - The code that represents composition information about a chemical substance that is not clearly identifiable as a single, specific chemical compound. (DE 5816, 1-character)
- Program System Chemical Substance Name - The name assigned to a chemical

substance by the EPA program that is interested in the substance. (DE 5810, 240-character)

- Program System Chemical Identifier Code - The identification code assigned to a chemical substance by the EPA program that is interested in the substance. (DE 5808, 12-character)
- Molecular Formula Code - A chemical formula showing the number of atoms of each element in a molecule of a chemical substance. (DE 5812, 20-character)
- Molecular Weight Quantity - The sum of the atomic weight of constituent atoms in a molecule of a chemical substance. (DE 5814, 10-numeric)

As required by the Agency *EPA Order - Chemical Abstract Service Registry Number Data Standard*, the data element DE 6165 is proposed as a mandatory data element for identification of a chemical substance.

## **APPENDIX A**

### **Proposed Standard Data Elements for Chemical Substances**

(See EDR for list of Data Elements for Chemical Substances)